

CERTIFICATE IN COMPUTATIONAL LINGUISTICS

Program Scope

The Certificate in Computational Linguistics is designed to provide academic training in the study of computational approaches to language analysis. The curriculum assumes no prior linguistic or programming knowledge and introduces students to a variety of computational methods and their theoretical underpinnings including: writing programs in Python to process raw texts (tokenization), discovering statistical patterns in linguistic data (frequency distribution), performing part-of-speech tagging, text segmentation, and classification (context-free grammars, dependency grammars), extracting meaning from texts, and applying various machine learning methods to data mining.

Program Learning Outcomes

1. Students will learn to identify grammatical categories and basic principles of phonological and syntactic grammar.
2. Students will learn to write programs in a programming language, e.g., Python, and to process raw texts.
3. Students will learn to discover statistical patterns in linguistic data, identify frequency distributions, and perform tokenization.
4. Students will learn to perform part-of-speech tagging, text segmentation, and classification.
5. Students will learn to build dependency grammar and extract meaning from texts.
6. Students will learn to apply various machine learning methods to data mining.

Certificate in Computation Linguistics – 15 units

Code	Title	Units
ENG 420	Introduction to the Study of Language	3
ENG 421	Syntax	3
ENG 620	Introduction to Computational Linguistics	3
ENG 680	Applied Computational Linguistics	3
Select One:		3
ENG 424	Phonology and Morphology	
CSC 620	Natural Language Technologies	